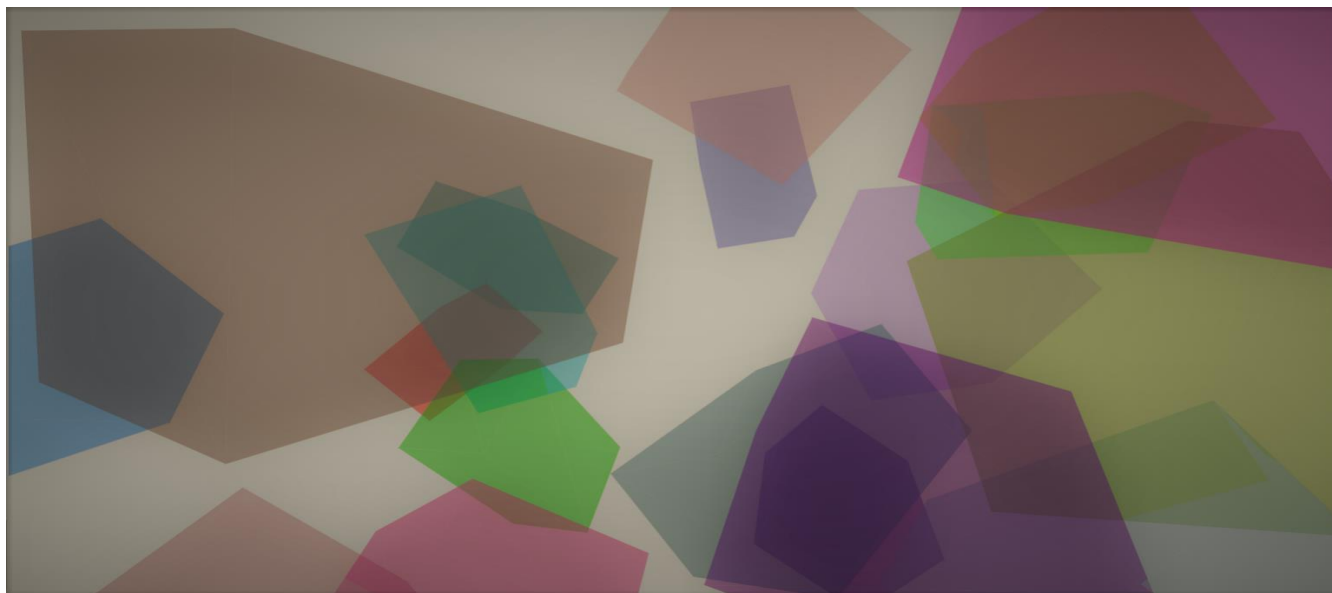




АНАЛИЗ НА МОДЕЛИ ЗА СКОРИНГ И РЕЙТИНГ ЧРЕЗ НЕВРОННА МРЕЖА

д-р Анатолий Антонов, д-р Венцислав Николов
31 март 2017

Еврориск Системи ООД
Ул. Генерал Киселов 31
9002 Варна, България
Тел. +359 52 612 367
Факс +359 52 612 371
info@eurorisksystems.com
www.eurorisksystems.com

Резюме: В настоящата статия се разглежда подход за анализ, съгласно регулаторните изисквания, на установена система за определяне на кредитен рейтинг и скоринг. За целта се използва модел с невронна мрежа, върху която се пренася реализираната логика и според свойствата на модела се определят чувствителности, значимост, независимост и други параметри на входните фактори.

Ключови думи: кредитен рейтинг, скоринг, регулаторни изисквания, анализ на фактори.

Кредитният скоринг представлява оценка за платежоспособността на кредитоискатели в бъдещи периоди. Тя се представя числено и от нея впоследствие за формиране на кредитен рейтинг могат да се обособят рейтингови степени по определена скала, които да се използват във вътрешни или регулаторни модели за оценка на кредитни рискове.

За получаване на кредитния скоринг се използват статистически методи за извличане на информация и зависимости от първични данни за кредитоискателите известни към момента на оценката. Системите за оценка на скоринг и рейтинг служат за прогнозиране при какви условия кредитоискателите ще бъдат в състояние да изплащат кредитни задължения и при какви условия те ще просрочат плащанията си или ще изпаднат в неизпълнение.

Данните, които се използват за такива оценки включват например: възраст, местожителство, нетни доходи, имущество, сума и срок на изплащане на кредита и др. Тези данни могат да бъдат анализирани с помощта на математически и статистически модели. Колкото по-пълна е историческата информация, толкова по-точни биха били построените модели. Системите за оценка на скоринг и рейтинг всъщност представляват класификатори от вида: към кой скоринг или рейтингова степен следва да се причисли даден кредитоискател с неговите първични данни. За реализация на тази цел се използват различни подходи, като йерархична система от правила, алгебрични изрази, таблици на решенията, графи и др.

В настоящия материал логиката на дадена скоринговата система се третира като постановена. Целта е да се представят методи за анализ, определяне на свойства и валидиране на логиката, на данните и на кредитен скоринг с помощта на модели основани на невронни мрежи. Принципът на провеждане на анализите е както следва:

- Логиката на скоринговия модел се пренася върху невронна мрежа:
Множество от скорирани кредитоискатели с техните входни данни и определени скоринги се третират като примери от предметната област на скорирането и се използват за обучение на невронна мрежа с приета структура;
- Невронната мрежа се анализира вместо оригиналната скорингова система:
След обучението невронната мрежата е натрупала стандартизирани знания във вид на вътрешни обекти и тегловни връзки между тях и представлява алтернативен стандартизиран модел, който позволява провеждането на анализи.
- Анализ и оценка на свойствата на скоринговия модел:
Анализите на свойствата на скоринговия модел се изискват регулаторно от надзорните органи съгласно РЕГЛАМЕНТ (ЕС) № 575/2013 и се отнасят до валидиране на скоринговия метод, определяне на значимост и независимост на факторите, определяне

на приноси и чувствителности на факторите, ортогонализация на скоринговата структура и др. Тези свойства формират надеждността на скоринговия модел като най-важен елемент при оценката на кредитните рискове.

1. Основни концепции на невронните мрежи

Невронните мрежи се дефинират като статистически модели за обработка на информацията. Те биват различни видове, като тези, използвани за скоринг системите спадат към надзирателно-обучаваните невронни мрежи, които работят в две фази. В първата фаза се извършват вътрешни настройки и калибриране в зависимост от предоставените входни и изходни данни (обучаващото множество от обучаващи примери за невронната мрежа). Всеки пример се състои от входни и съответни изходни резултати. Всички примери многократно се разпространяват през невронната мрежа, при което итеративно се извършват промени в изчислителната структура, някои връзки се усилват, други се отслабват. В разглеждания случай входните данни са данните на кредитополучателите, а изходните резултати представляват стойността на скоринга. Във втората фаза се обработват входни данни, за които се генерират предварително неизвестни резултати. На входа на вече обучената невронна мрежа се предоставят входни данни, за които не съществуват налични изходни резултати. Мрежата генерира изходни резултати, използвайки вече настроената вътрешна структура.

Изкуствените невронни мрежи са модели на естествените невронни мрежи и притежават някои подобни на тях характеристики. Обучението при естествените невронни мрежи представлява множество от физико-химични промени в синаптичните връзки между невроните с пространствена и времева активност. Информацията, която се акумулира в резултат от обучението се съхранява в синапсите под формата на концентрирани химични субстанции. Някои от основните характеристики на тези модели са представени по-долу.

- Обучение и адаптация
Фазата на обучение трябва да предхожда фазата на същинската работа. Обучението е по-сложен и времеотнемащ процес, при който се използва определен алгоритъм за намиране на правилните стойности за сила на връзките между отделните елементи. Всеки от представените обучаващи примери променя тези връзки в посока определена от зависимостите съдържащи се в примера. Възможно е след завършване на обучението, на някакъв по-късен етап да възникнат нови примери, така че да се премине към следващо ниво на обучение. Това обаче може да доведе до „забравяне“ на старите примери, ако те не бъдат предоставяни. Този ефект е полезен и може да се използва за реализация на еволюция на моделите по оста на времето. Обучаващото множество може да се разглежда като статистическа извадка от някаква безкрайна или много голяма генерална съвкупност. Така колкото е по-представителна извадката, толкова по-адекватно ще бъде обучението. То от своя страна може да се разглежда като оптимизационен процес, при който системата се настройва към представената извадка.
- Генерализация
Генерализацията е характеристика определяща възможността за вземане на адекватни решения на базата на предишен опит, приложен в нови ситуации. Процесът на възпроизвеждане на нова информация се основава на принципа, че подобните стимули

генерират подобни реакции. Колкото е по-голям предишният опит, толкова е по-вероятно невронната мрежа да генерира адекватен отговор в нова непозната ситуация, но това изисква и по-голяма мощност на невронната мрежа, което понякога усложнява значително нейната структура и я превръща в труден за разбиране модел подобен на черна кутия. Съществува и опасност тя да се адаптира прекалено добре само към представените примери (да се преобучи), но резултатите ѝ при генериране на нови данни да са значително по-слаби. Затова понякога част от обучаващото множество се отделя във валидиращо множество, с което се проверява способността на невронната мрежа да генерира адекватен отговор за непознати нови данни.

- Пространствена обработка на информацията

Принципът на работа на невронните мрежи, както във фазата на обучение, така и във фазата на генериране на нова информация, се базира на математически методи и решаване на векторно-матрични уравнения. В този процес някои части, области или отделни елементи на структурата се обособяват като реагиращи на определени стимули и могат да изпълняват ролята на представители за определени концепции в зависимост от поставената задача. Те отговарят на определени шаблони, характеристики, функции, задачи или цели. След обучението, структурата не променя връзките си и те могат да бъдат анализирани, а някои от тях оказващи несъществено влияние могат да бъдат и премахвани.

- Надеждност

Премахването на част от изчислителните елементи (неврони) често не повлиява съществено на работата на модела. Това дава възможност за изключване на част от връзките и невроните с несъществено влияние, което опростява модела и улеснява неговия анализ. Това е важно в случай, че се търсят научените концепции и се търсят заключения за силата на влиянието на факторите в представените примери. По този начин се опростява представянето и някои скрити зависимости стават очевидни.

Представените анализи в настоящия материал се базират на описаните по-горе основни характеристики.

2. Експериментална постановка за моделиране

Наличните експериментални данни за анализ на генерирания кредитен скоринг включват множество фактори, представящи информация за кредитоискатели като имена, възраст, заплата, спестявания, други доходи, нетни приходи, стойност на недвижима собственост, площ на недвижима собственост, брой хора в съжителство, образование, професия и др. Част от факторите не участват в определянето на скоринг-резултата, поради една или повече от следните причини:

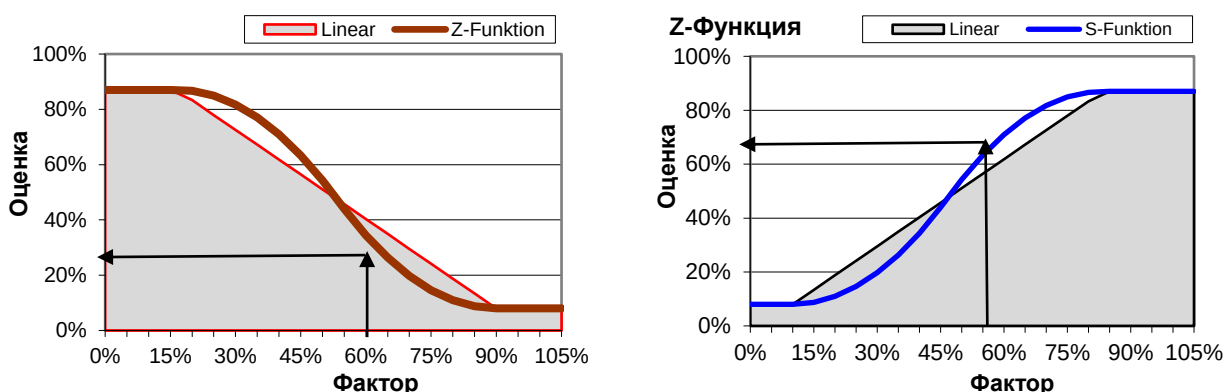
- Не носят полезна информация за модела.

Например имената на кредитоискателя не оказват влияние за работата на модела.

- Не могат да се представят в цифров вид.
Ако данните са представени като низове или друг тип данни, те трябва да се трансформират по определен начин (виж по-долу и калибриране) в числови стойности. Моделът работи само с числово представена информация и ако някои фактори не могат да се представят смислено в числов вид, те не могат да се включат като данни за обработване в него.
- Липса на достатъчно данни.
Например за фактора “други доходи за кредитоискателя” наличните ненулеви данни са около 5%, останалите са нули и не носят полезна информация за модела.
- Някои фактори се включват в други.
Например заплатата на кредитоискателя се включва в неговите нетни приходи.

Калибриране на входните фактори

Въздействията на входните фактори следва да се калибрират преди да се подадат на невронната мрежа с помощта на линейни или нелинейни (включващи насищане) S и Z функции (Фиг. 1).



Фиг. 1 Калибриране на входните фактори (S и Z функции)

Функциите S и Z трансформират входните фактори в калибрани въздействия от 0% до 100 % по следния начин:

S-Function: $Y = S(X, a, b)$; a, b – константи в зависимост от областта на фактора

за $X \leq a$: $S(X, a, b) = 0$;

за $X > b$: $S(X, a, b) = 1$

$$\text{за } a < X \leq (a+b)/2: \quad S(X,a,b) = 2(X-a)/(b-a);$$

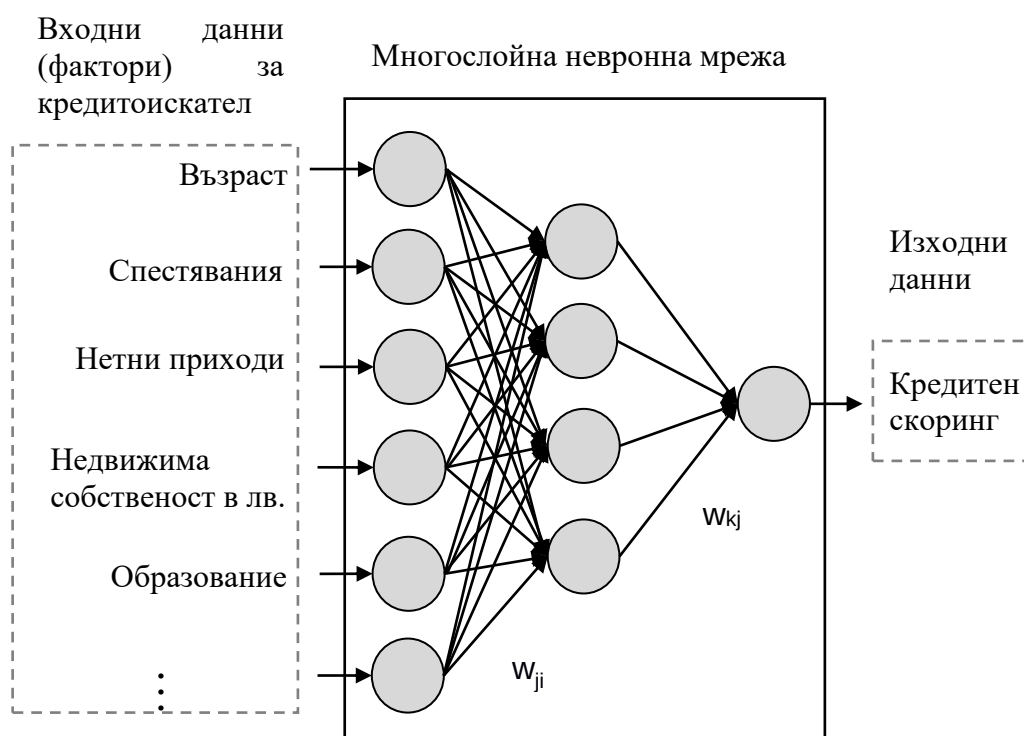
$$\text{за } (a+b)/2 < X \leq b: \quad S(X,a,b) = 1-2(b-X)/(b-a)$$

Z-Function: $Y = Z(X,a,b)$

$$Z(X,a,b) = 1 - S(X,a,b)$$

Структура на невронната мрежа

След подходящ подбор и обработка на факторите за конкретния случай се формират 22 фактора, които участват в построяването на мрежата. Невронната мрежа се обучава с няколко стотин записа с данни за кредитополучатели. Приетата структура на мрежата (Фиг. 2) е от тип многослоен персептрон с брой входове равен на броя фактори и един изход за скоринг стойността. Архитектурата включва и един скрит слой с размер определен от броя на обучаващите примери. Връзките между слоевете (матриците W_{ji} и W_{kj}) се представят в матричен вид чрез техните тегла.



Фиг. 2 Структура на многослойна невронна мрежа за определяне на кредитен скоринг

Агрегация на оценки в структурата

Агрегацията на групи от входни фактори или от частични оценки се извършва към обобщаваща оценка, като тя може да бъде линейна агрегация на претеглени фактори или нелинейна агрегация с насищане (Фиг. 3: a, b, c, ... са тегловни коефициенти) както следва:

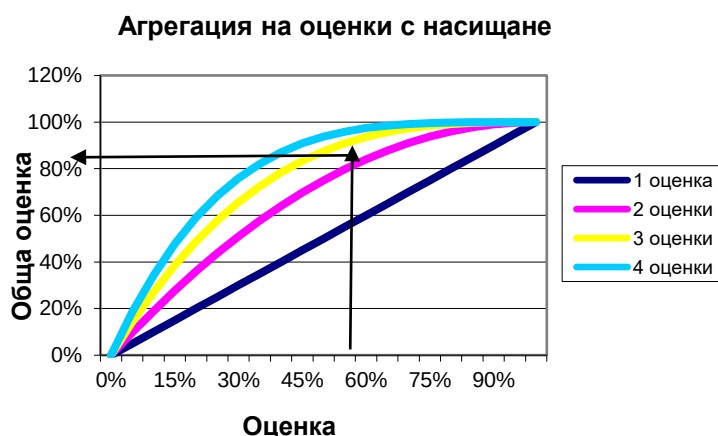
Линейна: Агрегация = $(a \cdot \text{Фактор1} + b \cdot \text{Фактор2} + \dots + N \cdot \text{ФакторN}) / (a + b + \dots + N)$

Пример: Агрегация = $(3,5 \cdot 63\% + 4,5 \cdot 36\% + 7,2 \cdot 48\%) / (3,5 + 4,5 + 7,2) = 47,9\%$

Насищане: Агрегация = $1 - (1 - \text{Фактор1}) \cdot (1 - \text{Фактор2}) \cdot \dots \cdot (1 - \text{ФакторN})$

Пример: Агрегация = $100\% - (100\% - 63\%) \cdot (100\% - 36\%) \cdot (100\% - 48\%) = 87,7\%$

Агрегациите се калибрират отново към 100% и не надвишават 100%.



Фиг. 3 Агрегация на оценки с насищане

Генериране на резултат от невронната мрежа

Обучението на невронната мрежа е сложен итеративен процес на настройка на тегловните матрици W_{ji} и W_{kj} от фиг. 2. Работата на всеки от елементите в невронната мрежа се състои от две стъпки: изчисление на сумарния претеглен входен сигнал:

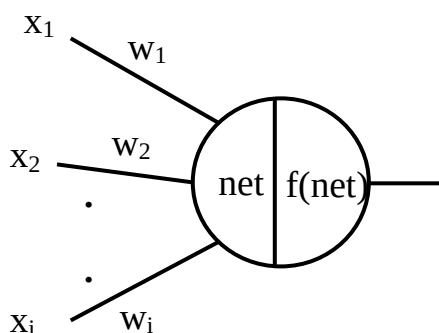
$$net_j = \sum_{i=1}^A w_{ji} x_i \quad (1)$$

и изчисление на активационното равнище:

$$h_j = f(net_j) \quad (2)$$

където $f()$ е активационната нелинейна функция на възлите на невронната мрежа, която тук е приета да бъде биполярен сигмоид (подобна на Z функцията от по-горе):

$$f(net_j) = -1 + \frac{2}{1+e^{-net_j}} = \frac{1-e^{-net_j}}{1+e^{-net_j}} \quad (3)$$



Фиг. 4 Изчислителен елемент (неврон) в невронната мрежа

Изразено в матрична форма разпространението на стойностите в невронната мрежа и генерирането на изход се състои от следните стъпки:

- 1) Изчисление на матрицата от сумарни претеглени входни стойности за елементите от междинния слой:

$$N_{ji} = W_{ji}X \quad (4)$$

- 2) Изчисление на матрицата от активационните равнища на елементите от междинния слой:

$$H_{ji} = f(N_{ji}) \quad (5)$$

- 3) Изчисление на матрицата от сумарни претеглени входни стойности за елементите от изходния слой:

$$N_{kj} = W_{kj}H_{ji} \quad (6)$$

- 4) Изчисление на изхода като активационните равнища на елементите от изходния слой:

$$S_{kj} = f(N_{kj}) \quad (7)$$

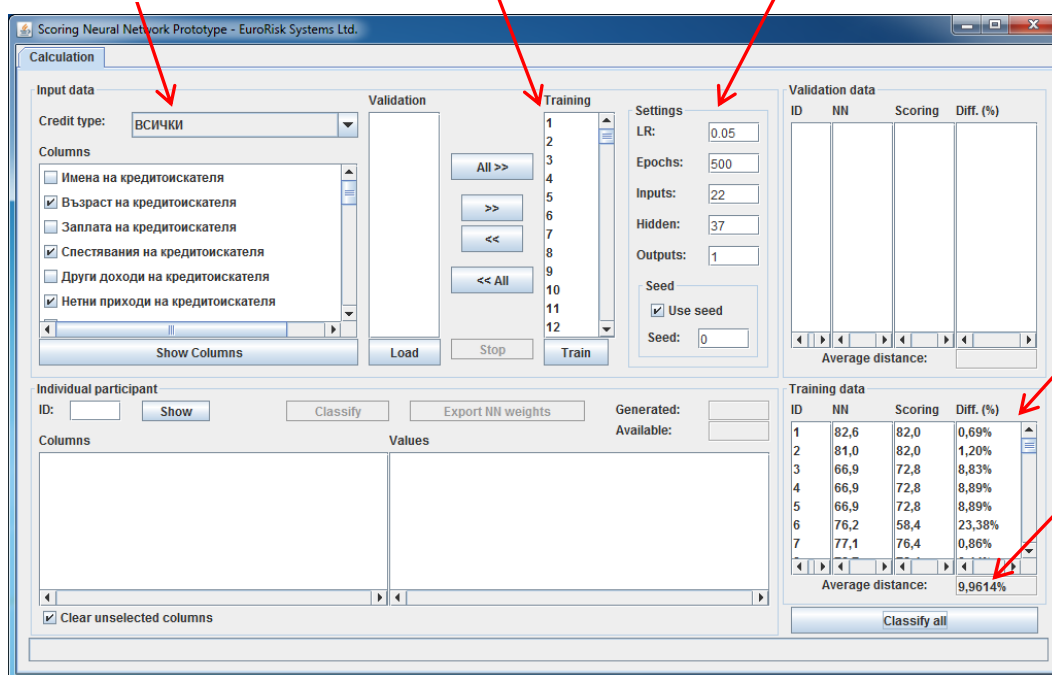
Елементите от изходния слой в случая се състоят само от един елемент, този който продуцира скоринг стойността.

Обучение с използване на всички обучаващи примери

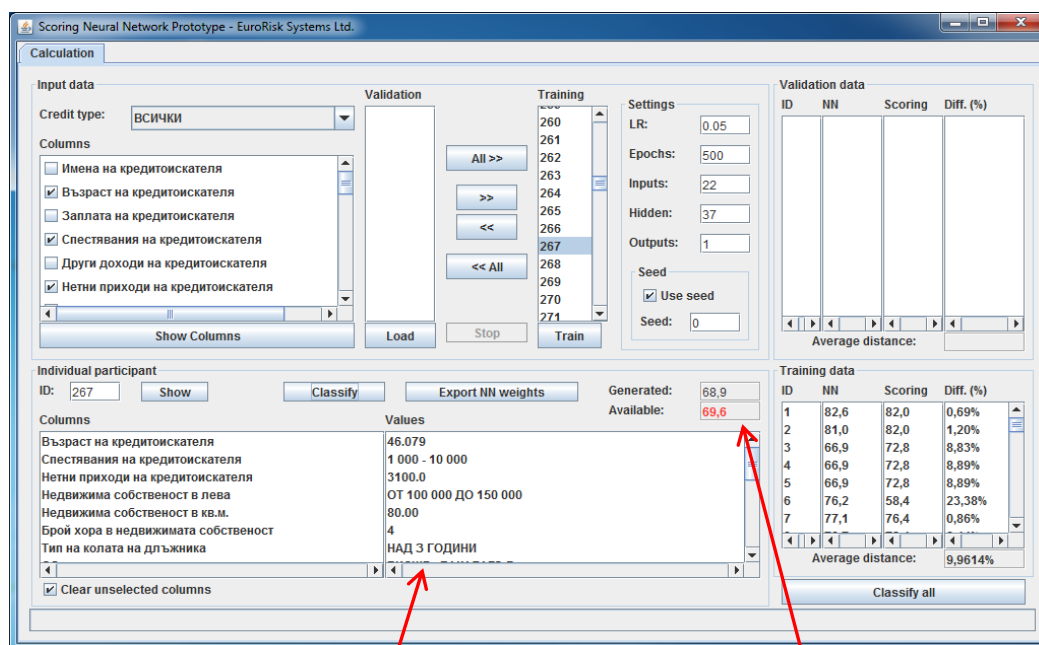
Един обучаващ пример за невронната мрежа представлява един запис с данните за кредитополучател с известен скоринг. Скорингът се приема за еталонен и трябва да е бил коректно определен. Моделът на невронната мрежа може да се разглежда като сложна предварително неизвестна функция с много променливи. В резултат от обучението тази функция се конкретизира и в следващата фаза тя може да се използва за получаване на резултат от нея при подаване на конкретни фактически аргументи, представляващи данните за нов кредитополучател, чиито кредитен скоринг предстои да бъде определен. Поради тази причина е важно обучението да се извърши по възможно най-добрия начин.

За експерименталните проверки на работата на модела за валидация и генериране на скоринг е разработен прототип на модел с невронна мрежа, чиито резултати при използване на всички обучаващи примери е показан на фиг. 5.

Тип кредит Всички кредитополучатели са избрани за обучение на мрежата Автоматично определени параметри на мрежата



Фиг. 5 Обучение с използване на всички данни за кредитополучатели



Фиг. 6 Стойност на скоринга за индивидуален кредитополучател

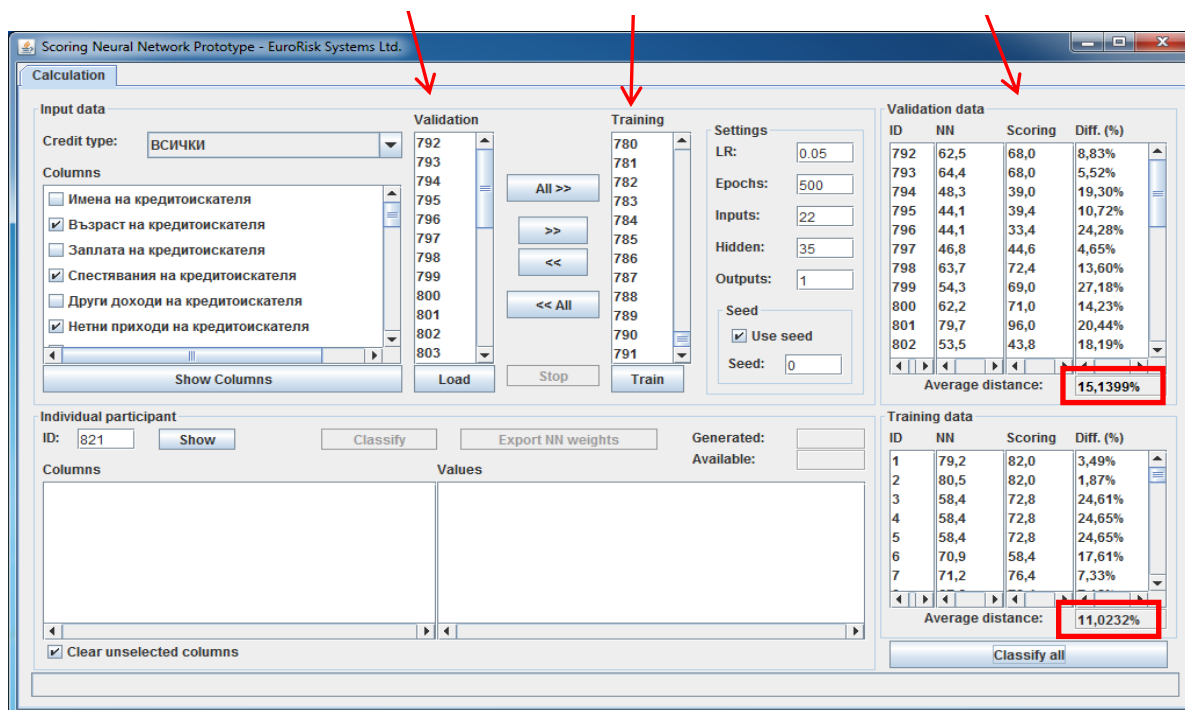
Резултати от обучение с използване на валидиращо подмножество

Преди фазата на обучението част от наличните обучаващи примери се отделят като валидираща извадка, която не участва в обучението. С нея се проверява адекватността на модела за нови непознати данни след края на обучението. Резултатите за обучаващото и валидиращото множество са показани на фиг. 7. Вижда се, че за валидиращото множество средното отклонение е 15.14%, което е малко по-голямо от това за обучаващото множество 11.02%, тъй като обучаващото множество е известно за модела.

Част от данните са
отделени за валидация

Обучаващо
множество

Резултати за
валидиращото множество



Фиг. 7 Скоринг резултати при използване на обучаващо и валидиращо множества

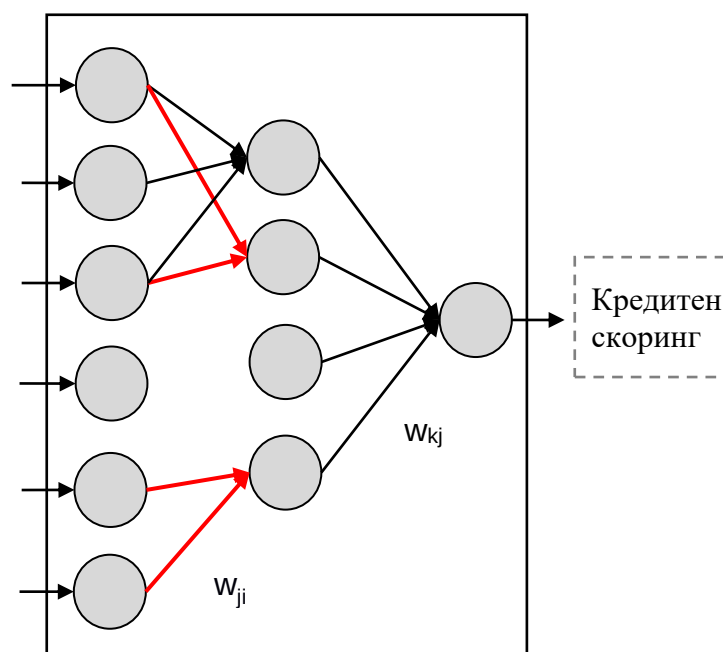
3. Анализи на модела

Анализ и редукция на структурата на модела

След обучението на невронната мрежа тегловните коефициенти към невроните се запазват и могат да се използват за индивидуални изчисления с произволен входен вектор от данните на кредитополучател и за анализ на мрежата. С цел намаляване на сложността на модела, определяне на значимост на факторите и редуциране на броя на невроните, теглата, които са с ниски стойности, могат да се използват за премахване на незначимите връзки в

мрежата. В резултат е възможно цялостно елиминиране на незначими фактори или вътрешни неврони, което води до „избистряне“ на мрежата и позволява както по-лесното интерпретиране и разбиране на модела, така и откриването на закодирани в тегловната матрица зависимости. Някои от възлите на невронната мрежа (редовете на тегловната матрица) научават определени концепции и те могат да бъдат открити и интерпретирани чрез такова редуциране. Освен това някои от факторите (колони на тегловната матрица) оказват по-голяма тежест при определяне на скоринг стойността и с редуцирането те по-ясно се открояват спрямо останалите фактори. Тези възможности причисляват анализа към групата на методи за откриване на знания (Data Mining). Редуцирането в експерименталните изследвания е извършено само за тегловната матрица между входния и междинния (скрит) слой на невронната мрежа. На фиг. 8 е представена опростена структура на мрежата след премахване на незначимите връзки. Забелязват се следните особености:

- Първи, втори и четвърти неврон от скрития слой зависят вече само от прегледен брой входни фактори, четвърти неврон е независим от другите неврони. Тези неврони представят автоматично определени междинни понятия в предметната област с ясна интерпретация относно входните фактори. Например, ако входните фактори: заплата, наеми, други доходи се отнасят към един и същ вътрешен неврон, то лесно може да се заключи, че невронът обобщава понятието „приходи“.
- Четвъртият входен фактор и третият неврон могат да се премахнат като неоказващи влияние.



Фиг. 8 Редуцирана структура на невронната мрежа

За редуциране на връзките в тегловната матрица предварително се задава прагова стойност t на стандартното отклонение e_{ave} за всички M обучаващи примери между скоринга

S при редуцирана и скоринга R при нередуцирана тегловна матрица W_{ji} . Извършва се линейно търсене на стойност за r , такава, че e_{ave} да не надвишава t .

$$e_{ave} = \sqrt{\frac{1}{M-1} \sum_{m=1}^M (D_m - \bar{D}_m)^2} \quad (8)$$

$$D_m = \frac{(S_m - R_m)}{S_m} \quad (9)$$

Където S_m е получен от (7) за m -ти кредитополучател, а R_m е получен по същия начин, но с редуцирана матрица W_{ji} в (4) както следва:

$$w_{ji} = \begin{cases} w_{ji}, & \text{ако } w_{ji} < r \\ 0, & \text{ако } w_{ji} \geq r \end{cases} \quad (10)$$

Резултатите от търсенето на r са показани на фиг. 9, където се вижда част от редуцираната тегловна матрица (пълната матрица включва 22 фактора x 37 неврона във вътрешния слой). Под нея са изведени следните параметри или резултати:

Limit	r (най-нисък тегловен коефициент за редуцирането)
Scoring0	скоринг без редукция на тегловната матрица;
Error	t (максимално стандартно отклонение между S и R)
Delta	стъпка на линейното търсене;
Fill	запълване на матрицата (процент на нередуцираните връзки от тегловната матрица към всички връзки)

Intermediate Nodes	Възраст на кредитоискателя	Спестявания на кредитоискателя	Нетни приходи на кредитоискателя	Недвижима собственост в лева	Недвижима собственост в кв. м.	Брой хора в недвижимата собственост	Тип на колата на длъжника	Образование на длъжника	Семеен статус на кредитоискателя	Тип предприятие на кредитоискателя	Сектор в който работи кредитоискателя
unit 27											
unit 28										0,231	0,296
unit 29			-0,229								
unit 30											
unit 31			-0,270								
unit 32	0,631	0,844	0,803	1,084	-0,793		-0,254			-0,335	-0,851
unit 33											
unit 34											
unit 35										-0,216	
unit 36											
unit 37											

NN Results					Weights					
Scoring	Min	Max	Scale	Shift	Limit	Scoring0	Diff	Error	Delta	Fill
69,058	14	96	0,02195	1,20732	21,00%	69,0583		10,00%	1,00%	7,99%

Фиг. 9 Редуциране на тегловна матрица

В резултат от редуцирането на теглата се забелязва открояването на определени неврони от вътрешния слой (напр. unit 32), които приемат по-силно информацията от входните фактори и други (напр. unit 30, unit 33 и т.н.), които не влияят при определянето на скоринг стойността и могат да се премахнат. По подобен начин някои от факторите се открояват спрямо други, което показва тяхната по-голяма значимост за скоринга. Например нетните приходи на кредитоискателя оказват съществено влияние върху скоринга, за разлика от броя хора в съжителство или семейния статус на кредитоискателя, които влияят по-незначително (в рамките на цялата мрежа). Нередуцираната част от матрицата е представена на Фиг. 10.

Intermediate Nodes	Възраст на кредитоискателя	Спестявания на кредитоискателя	Нетни приходи на кредитоискателя	Недвижима собственост в лева	Недвижима собственост в кв. м.	Брой хора в недвижимата собственост	Тип на колата на длъжника	Образование на длъжника	Семеен статус на кредитоискателя	Тип предприятие на кредитоискателя	Сектор в който работи кредитоискателя
unit 28	-0,009	-0,164	0,126	-0,058	0,115	-0,001	0,077	0,081	0,148	0,231	0,296
unit 29	0,119	0,050	-0,229	0,033	-0,125	0,018	-0,177	-0,152	-0,141	-0,049	-0,149
unit 30	-0,203	-0,042	0,142	0,066	0,073	-0,032	0,073	0,072	0,095	0,104	0,136
unit 31	0,124	-0,075	-0,270	-0,111	-0,183	0,080	-0,138	-0,012	0,021	-0,121	-0,100
unit 32	0,631	0,844	0,803	1,084	-0,793	0,206	-0,254	0,052	-0,110	-0,335	-0,851
unit 33	-0,035	0,036	0,149	-0,028	0,060	-0,070	0,068	0,000	-0,018	0,129	0,149
unit 34	-0,090	0,088	0,085	-0,046	-0,070	0,005	0,060	-0,013	0,044	-0,110	-0,065
unit 35	0,057	0,006	-0,127	-0,027	-0,140	0,081	-0,053	-0,028	-0,067	-0,216	-0,139
unit 36	-0,058	-0,008	-0,079	-0,053	-0,150	-0,001	0,034	-0,085	-0,056	-0,138	-0,122
unit 37	0,004	0,014	-0,004	0,058	-0,060	-0,035	0,075	-0,028	0,026	0,014	0,047

Фиг. 10 Нередуцирана тегловна матрица

Обобщеният ефект от редукцията на структурата на мрежата е следният:

- намаляване на сложността на модела;
- определяне на значимост и принос на факторите;
- оптимизация, цялостно елиминиране на незначими фактори или вътрешни неврони;
- „избистряне“ на мрежата, по-лесното интерпретиране и разбиране на модела;
- откриване на закодирани в тегловната матрица зависимости (data mining).

Принос и чувствителности на фактори

При наличие на тегловната матрица чрез която се генерират скоринг стойностите, може да се определи чувствителността на скоринг резултата спрямо всеки от факторите, т.е. при промяна на даден фактор с колко се променя скоринг резултатът.

$$c_i = \frac{\partial S}{\partial f_i} \quad (11)$$

По този начин може да се определи и независимост между факторите като се сравнява сумата от чувствителностите на всички фактори с чувствителността на всички фактори при едновременна промяна.

$$c_a = \frac{\partial^n S}{\partial f_1 \partial f_2 \dots \partial f_n} \text{ и сравнение със } c_b = \sum \frac{\partial S}{\partial f_i} \quad (12)$$

Определянето на независимост на факторите се базира на Тейлорово развитие, където всички по-висши или смесени диференциали са нули. При наличие на независимост чувствителностите могат да се интерпретират като линейни приноси на факторите към скоринговия резултат.

За извършване на изчисленията, предварително се задава промяна Δc и всички стойности w_{ji} в колоната на тегловната матрица W_{ji} съответстваща на фактор i (теглата на всички връзки свързани към входен неврон i) се променя по следния начин:

$$w_{ji} = w_{ji}(1 - \Delta c_i) \quad (13)$$

след което се изчислява скоринг стойността S от (7).

Примерни фактори	Чувствителност
Възраст на кредитоискателя	-0,33%
Спестявания на кредитоискателя	0,68%
Нетни приходи на кредитоискателя	13,58%
Недвижима собственост в лева	-0,46%
Недвижима собственост в кв.м.	4,07%
Брой хора в недвижимата собственост	1,11%
Тип на колата на длъжника	-0,15%

Образование на длъжника	0,73%
Семеен статус на кредитоискателя	0,69%
Тип предприятие на кредитоискателя	5,57%
Сектор в който работи кредитоискателя	-0,02%
Тип длъжност	-0,20%
Източник на приходите на кредитоискателя	-1,06%
Общ стаж на кредитоискателя	3,91%
Длъжност в банката	2,31%
Област на кредитополучателя	0,18%
Целева група на банката	0,00%
Тип на кредита	1,47%
Размер на кредита	6,79%
Лихва	-15,72%
Валута на кредита	-0,28%
Срок на погасяване на кредита в месеци	-5,82%
Сума на чувствителностите	17,045%
Чувствителност при едновременна промяна	17,058%
Разлика, независимост и линейност	0,0013%

Таблица 1. Чувствителности на скоринга спрямо факторите

Чувствителността c_i се получава като процентна разлика между пресметнатия по този начин S и скоринг стойността S_0 без промяна на фактор i . Аналогично се изчислява S след едновременна промяна на всички фактори и се пресмята процентната разлика с S_0 . Резултатите са показани в Таблица 1.

Стойностите в Таблица 1 са дадени за $\Delta c_i = 1\%$ за всеки от факторите, изчислено е влиянието, което оказва тази промяна върху скоринг стойността спрямо непроменената стойност на фактора. Чувствителностите спрямо скоринга се изчисляват като сума от всяка промяна на фактор, както и от едновременните промени на всички фактори. При резултатите показани в Таблица 1 се вижда, че сумарната стойност на промените е приблизително равна на стойността при едновременна промяна, което определя нулеви влияния на всички по-висши и смесени промени, откъдето следва независимостта и линейността на факторите.

Обобщеният ефект от изчислението на чувствителностите спрямо факторите е следният:

- Определяне на значимост на факторите при промени = факторен риск на скоринга;
- Определяне на риска при неблагоприятни едновременно промени = риск на скоринга;
- Проверка за независимост на факторите;
- Определяне на линейност и принос на факторите.

4. Бъдещо развитие

Представената методика има за цел анализ и валидация на системи на кредитен скоринг, с които се произвеждат скоринги и кредитни рейтинги от заинтересовани кредитни институции. Валидацията и анализът се извършват на базата на предварително представена извадка от примери за скорирание, с които се обучава невронна мрежа в качеството си на аналитичен модел. Анализът на такъв модел може да бъде развит в следните посоки:

- Еволюция на скоринговия модел:
Обучение на невронната мрежа с нови примери, които се променят с времето и еволюират, което обуславя структурни промени в мрежата и промяна на тегловните коефициенти. Тук от интерес е автоматичното извличане на промените в мрежата, които да се реализират чрез промени в логиката и теглата на скоринговата система;
- Автоматично определяне на калибрирането на входните фактори при зададен диапазон и възможни стойности, напр. опит на управителя = (отличен, добър, среден...);
- При наличие на регулярни скорингови примери по оста на времето, т.е. периодично скорирание на длъжниците, напр. през 3 месеца, са възможни следните анализи:

- Тест за стационарност: постоянни статистики по оста на времето като очаквана стойност, тренд и вариация;
- Тест за сезонност: повтарящи се промени във входните данни и в поведението на рейтинговата система, напр. през зимните месеци или при политически събития. Сезонността се отличава от цикличността, която може да има промени с различна продължителност. Сезонността е прогнозируема;
- Тест за корелация между факторите и корелация с други стопански фактори: Корелацията между две серии се пресмята както следва:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2]}} \quad (14)$$

- Тест за автокорелация на факторите = корелацията между точки от дадена серия отстоящи на еднакви разстояния. При дадени N наблюдения $x_1, \dots, x_N$, могат да се образуват N-1 двойки от наблюдения $(x_1, x_2), (x_2, x_3), \dots, (x_{N-1}, x_N)$. Разглеждайки първите стойности на всяка двойка като една променлива, а вторите стойности на двойките като втора променлива, може да се изчисли корелационният

коэффициент между x_t и x_{t+1} :

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)}) (x_{t+1} - \bar{x}_{(2)})}{\sqrt{\left[\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})^2 \sum_{t=1}^{N-1} (x_{t+1} - \bar{x}_{(2)})^2 \right]}} \quad (15)$$

По същия начин могат да се намерят корелациите между наблюдения, отстоящи на разстояние k едно от друго:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x}) (x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad (16)$$

което се нарича автокорелационен коэффициент с лаг k . За по-малки стойности на N е по-точна формулата:

$$c_k = \frac{1}{N-k} \sum_{t=1}^{N-k} (x_t - \bar{x}) (x_{t+k} - \bar{x}) \quad (17)$$